

WHAT IS CLAIMED IS:

1. A method of recognizing a frame in a document image, comprising:
- inputting document image data;
- 5 extracting a black pixel rectangle that circumscribes continuous black pixels, the black pixel rectangle defining a parent rectangle;
- determining whether or not the parent rectangle is a frame candidate based upon a predetermined set of first criteria;
- extracting a white pixel rectangle that circumscribes continuous white pixels
- 10 within the parent rectangle of the frame candidate; and
- further determining whether or not the frame candidate is a frame based upon a predetermined set of second criteria, the second criteria including at least a comparison of a certain aspect between the white pixel rectangle and the black pixel rectangle.
- 15 2. The method of recognizing a frame in a document image according to claim 1 wherein the predetermined set of the first criteria includes a comparison of one side of the black pixel rectangle to a first predetermined threshold value.
3. The method of recognizing a frame in a document image according to claim 1 wherein
- 20 the predetermined set of the first criteria includes a comparison of a ratio of sides of the black pixel rectangle to a second predetermined threshold value.
4. The method of recognizing a frame in a document image according to claim 1 wherein the predetermined set of the first criteria includes a comparison of a black pixel density of
- 25 the black pixel rectangle to a third predetermined threshold value.
5. The method of recognizing a frame in a document image according to claim 1 wherein the predetermined set of the second criteria includes additional steps of:
- comparing a center difference in center position between the white pixel rectangle
- 30 and the black pixel rectangle to a fourth predetermined threshold value; and
- recognizing the frame candidate as a frame if the center difference is equal to or smaller than the fourth predetermined threshold value.

6. The method of recognizing a frame in a document image according to claim 1 wherein the predetermined set of the second criteria includes additional steps of:

5 determining a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle;

further determining a max-to-min difference between a maximal value of the pixel value differences and a minimal value of the pixel value differences;

comparing the max-to-min difference to a fifth predetermined threshold value; and

10 recognizing the frame candidate as a frame if the max-to-min difference is equal to or smaller than the fifth predetermined threshold value.

7. The method of recognizing a frame in a document image according to claim 1 wherein the predetermined set of the second criteria includes additional steps of:

15 determining a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle;

further determining a ruled line difference between each of the pixel value differences and a predetermined ruled line thickness;

20 comparing the ruled line difference to a sixth predetermined threshold value; and

recognizing the frame candidate as a frame if the ruled line difference is equal to or smaller than the sixth predetermined threshold value.

8. The method of recognizing a frame in a document image according to claim 7 wherein the predetermined set of the second criteria includes additional steps of:

25 determining whether or not any one of the pixel value differences is zero if the ruled line difference is larger than the sixth predetermined threshold value, a matched side being defined as a side for the pixel value difference being zero;

determining whether or not a ruled line exists on the matched side if the pixel value difference is zero; and

30 recognizing the frame candidate as a frame if the ruled line exists on the matched side.

9. The method of recognizing a frame in a document image according to claim 1 wherein the predetermined set of the second criteria includes additional steps of:

- generating a histograms of the length of the continuous black pixels;
- assuming a width of a ruled line based upon a median value in the histogram;
- 5 determining a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle;
- determining a line thickness difference between the pixel value difference and the assumed width of the ruled line;
- comparing the line thickness difference to a seventh predetermined threshold
- 10 value; and
- recognizing the frame candidate as a frame if the line thickness difference is equal to or smaller than the seventh predetermined threshold value.

10. The method of recognizing a frame in a document image according to claim 9 wherein the predetermined set of the second criteria includes additional steps of:

- determining whether or not any one of the pixel value differences is zero if the ruled line difference is larger than the sixth predetermined threshold value, a matched side being defined as a side for the pixel value difference being zero;
- determining whether or not a ruled line exists on the matched side if the pixel
- 20 value difference is zero; and
- recognizing the frame candidate as a frame if the ruled line exists on the matched side.

11. The method of recognizing a frame in a document image according to claim 1 wherein the predetermined set of the second criteria includes additional steps of:

- generating a histograms of the length of the continuous black pixels;
- assuming a width of a ruled line based upon a median value in the histogram;
- determining a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle, the pixel value
- 30 difference being defined as an eight threshold value;
- comparing a double width that is twice the width of the ruled line to the eight threshold value;

comparing the pixel value difference to eight threshold value plus five if the eight threshold value is smaller than the double width;

determining whether or not any one of the pixel value differences is zero if the pixel value difference is smaller than the eight threshold value plus five, a matched side
5 being defined as a side for the pixel value difference being zero;

determining whether or not a ruled line exists on the matched side if the pixel value difference is smaller than eight threshold value plus five; and

recognizing the frame candidate as a frame if the ruled line exists on the matched side.

10

12. The method of recognizing a frame in a document image according to claim 1 wherein the predetermined set of the second criteria includes additional steps of:

generating a histogram of the length of the continuous black pixels;

assuming a width of a ruled line based upon a median value in the histogram;

15 determining a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle, the pixel value difference being defined as an eight threshold value;

comparing a double width that is twice the width of the ruled line to the eight threshold value;

20 comparing the pixel value difference to eight threshold value plus five if the eight threshold value is smaller than the double width;

determining whether or not any one of the pixel value differences is zero if the pixel value difference is smaller than the eight threshold value plus five; and

25 recognizing the frame candidate as a frame if any one of the pixel value differences is not zero.

13. The method of recognizing a frame in a document image according to claim 1 further comprising:

30 integrating into a single area the frames as determined by the first and second criteria and the frame candidates as determined by the first criteria but not by the second criteria based upon a common orientation of the frames and the frame candidates;

selecting one of the frame candidates at a time; and

determining whether or not the selected one of the frame candidates is a frame based upon a predetermined set of third criteria.

14. A computer readable medium storing a computer program for recognizing a frame in a document image, the computer program providing computer instructions to perform the following steps of:

- inputting document image data;
- extracting a black pixel rectangle that circumscribes continuous black pixels, the black pixel rectangle defining a parent rectangle;
- 10 determining whether or not the parent rectangle is a frame candidate based upon a predetermined set of first criteria;
- extracting a white pixel rectangle that circumscribes continuous white pixels within the parent rectangle of the frame candidate; and
- further determining whether or not the frame candidate is a frame based upon a predetermined set of second criteria, the second criteria including at least a comparison of a certain aspect between the white pixel rectangle and the black pixel rectangle.

15. A system for recognizing a frame in a document image, comprising:
- 20 an input unit for inputting document image data;
 - a black pixel rectangle extraction unit connected to said input unit for extracting a black pixel rectangle that circumscribes continuous black pixels, the black pixel rectangle defining a parent rectangle;
 - a first determination unit connected to said black pixel rectangle extraction unit
 - 25 for determining whether or not the parent rectangle is a frame candidate based upon a predetermined set of first criteria;
 - a white pixel rectangle extraction unit connected to said first determination unit for extracting a white pixel rectangle that circumscribes continuous white pixels within the parent rectangle of the frame candidate; and
 - 30 a second determination unit connected to said black pixel rectangle extraction unit and said white pixel rectangle extraction unit for further determining whether or not the frame candidate is a frame based upon a predetermined set of second criteria, the second

criteria including at least a comparison of a certain aspect between the white pixel rectangle and the black pixel rectangle.

16. The system for recognizing a frame in a document image according to claim 15
5 wherein said first determination unit compares one side of the black pixel rectangle to a first predetermined threshold value.

17. The system for recognizing a frame in a document image according to claim 15
10 wherein said first determination unit compares a ratio of sides of the black pixel rectangle to a second predetermined threshold value.

18. The system for recognizing a frame in a document image according to claim 15
15 wherein said first determination unit compares a black pixel density of the black pixel rectangle to a third predetermined threshold value.

19. The system for recognizing a frame in a document image according to claim 15
20 wherein said second determination unit compares a center difference in center position between the white pixel rectangle and the black pixel rectangle to a fourth predetermined threshold value recognizes the frame candidate as a frame if the center difference is equal to or smaller than the fourth predetermined threshold value.

20. The system for recognizing a frame in a document image according to claim 15
25 wherein said second determination unit determines a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle, said second determination unit further determining a max-to-min difference between a maximal value of the pixel value differences and a minimal value of the pixel value differences, said second determination unit comparing the max-to-min difference to a fifth predetermined threshold value, and said second determination unit recognizing the frame candidate as a frame if the max-to-min difference is equal to or smaller than the fifth
30 predetermined threshold value.

21. The system for recognizing a frame in a document image according to claim 15 wherein said second determination unit determines a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle, said second determination unit further determining a ruled line difference
5 between each of the pixel value differences and a predetermined ruled line thickness, said second determination unit comparing the ruled line difference to a sixth predetermined threshold value, and said second determination unit recognizing the frame candidate as a frame if the ruled line difference is equal to or smaller than the sixth predetermined threshold value.
- 10 22. The system for recognizing a frame in a document image according to claim 21 wherein said second determination unit further determines whether or not any one of the pixel value differences is zero if the ruled line difference is larger than the sixth predetermined threshold value, a matched side being defined as a side for the pixel value
15 difference being zero, said second determination unit determining whether or not a ruled line exists on the matched side if the pixel value difference is zero, and said second determination unit recognizing the frame candidate as a frame if the ruled line exists on the matched side.
- 20 23. The system for recognizing a frame in a document image according to claim 15 wherein said second determination unit generates a histogram of the length of the continuous black pixels, said second determination unit assuming a width of a ruled line based upon a median value in the histogram, said second determination unit determining a
25 pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle, said second determination unit further determining a line thickness difference between the pixel value difference and the assumed width of the ruled line, said second determination unit comparing the line thickness difference to a seventh predetermined threshold value, and said second determination unit recognizing the frame candidate as a frame if the line thickness difference is equal to or
30 smaller than the seventh predetermined threshold value.

24. The system for recognizing a frame in a document image according to claim 23 wherein said second determination unit determines whether or not any one of the pixel value differences is zero if the ruled line difference is larger than the sixth predetermined threshold value, a matched side being defined as a side for the pixel value difference being zero, said second determination unit determining whether or not a ruled line exists on the matched side if the pixel value difference is zero, and said second determination unit recognizing the frame candidate as a frame if the ruled line exists on the matched side.

25. The system for recognizing a frame in a document image according to claim 15 wherein said second determination unit generates a histograms of the length of the continuous black pixels, said second determination unit assuming a width of a ruled line based upon a median value in the histogram, said second determination unit determining a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle, the pixel value difference being defined as an eight threshold value, said second determination unit comparing a double width that is twice the width of the ruled line to the eight threshold value, said second determination unit comparing the pixel value difference to eight threshold value plus five if the eight threshold value is smaller than the double width, said second determination unit determining whether or not any one of the pixel value differences is zero if the pixel value difference is smaller than the eight threshold value plus five, a matched side being defined as a side for the pixel value difference being zero, said second determination unit determining whether or not a ruled line exists on the matched side if the pixel value difference is smaller than eight threshold value plus five, and said second determination unit recognizing the frame candidate as a frame if the ruled line exists on the matched side.

26. The system for recognizing a frame in a document image according to claim 15 wherein said second determination unit generates a histograms of the length of the continuous black pixels, said second determination unit assuming a width of a ruled line based upon a median value in the histogram, said second determination unit determining a pixel value difference in coordinates of each of corresponding four corners between the white pixel rectangle and the black pixel rectangle, the pixel value difference being defined as an eight threshold value, said second determination unit comparing a double width that

is twice the width of the ruled line to the eight threshold value, said second determination unit comparing the pixel value difference to eight threshold value plus five if the eight threshold value is smaller than the double width, said second determination unit determining whether or not any one of the pixel value differences is zero if the pixel value difference is smaller than the eight threshold value plus five, and said second determination unit recognizing the frame candidate as a frame if any one of the pixel value differences is not zero.

27. The system for recognizing a frame in a document image according to claim 15 further comprising:

an imaginary continuous frame area generation unit connected to said first determination unit and said second determination unit for integrating into a single area the frames as determined by the first and second criteria and the frame candidates as determined by the first criteria but not by the second criteria based upon a common orientation of the frames and the frame candidates; and

a frame recognition processing unit connected to said imaginary continuous frame area generation unit for selecting one of the frame candidates at a time and determining whether or not the selected one of the frame candidates is a frame based upon a predetermined set of third criteria.